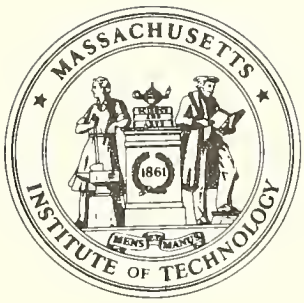

working paper
department
of economics

A CORRESPONDENCE PRINCIPLE FOR SIMULTANEOUS EQUATION MODELS

by

Franklin M. Fisher

Number 9 --- November 1967

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

A CORRESPONDENCE PRINCIPLE FOR SIMULTANEOUS EQUATION MODELS

by

Franklin M. Fisher

Number 9 --- November 1967

This paper was largely written during my tenure of a Ford Foundation Faculty Research Fellowship in Economics, while visiting the Hebrew University. The ideas here discussed grew out of discussions with Edwin Kuh and research carried on jointly with him. The views expressed in this paper are the author's sole responsibility, and do not reflect those of the Department of Economics, nor of the Massachusetts Institute of Technology.

1. Introduction

This paper presents a set of conditions which, it is argued, must be satisfied by any properly specified and estimated simultaneous equation model. Those conditions are derived by considering the implications of the view that simultaneous models are limiting approximations to non-simultaneous models in which certain time lags approach zero. The results lead immediately to a battery of easily applied tests which such a model must pass. Those tests have two important properties. First, they apply to the interaction of sets of equations rather than to single equations, and therefore test aspects of a model which are generally not tested in the specification and estimation of individual equations. Second, the tests apply both to the model as a whole and to any submodel formed by deletion of one or more equations and treatment of the corresponding endogenous variables as if they were predetermined. Since the functioning of submodels is tested, our results can be used to locate problems of specification.

A byproduct of our results is that any model (or submodel) satisfying our conditions lends itself readily to simulation and forecasting experiments with the values of the current endogenous variables generated by an iterative technique of extreme computational simplicity. This is an important feature in large models containing nonlinear identities or other nonlinearities. Nevertheless, the tests developed here are applicable to any simultaneous model, linear or nonlinear, and should be applied as a matter of standard econometric practice.

The theoretical development of our tests is followed by a discussion of their application to a revised version of the Klein-Goldberger model.

2. Simultaneity as a Limit

A well-known and highly convincing position on the nature of simultaneity in econometric models is that such models are only approximations to the true state of affairs. In fact, causation takes time, and the reactions given by the equations of the model truly occur not instantaneously but with a very small time lag. Unfortunately, however, data do not come to us sufficiently finely divided in time to allow us to observe such fast-moving reactions, so we take simultaneous instantaneously-holding relations as approximations, valid between the observations which nature allows us. Time lags are thus considered negligible provided they are sufficiently small.[1]

The consequences of this position for parameter estimation when observations occur at discrete points of time separated by an interval much larger than that in which the true reactions take place have been discussed in the literature.[2] In fact, however, this particular variant of the above view does not seem a very realistic one. We very seldom have observations on the value of a particular variable at precise discrete moments in time, and, if we do, we seldom use the observations in that form.[3] Much more common is the case in which the observations either by necessity or by choice are in the form of averages or sums over a non-zero time interval. Simultaneous

---

[1] This view has been discussed at length by Bentzel and Hansen [2]. The basic position on causation has been vigorously maintained by H. Wold in several works, e.g., Wold and Jureen [11].

[2] Strotz [10]; but see also Gorman [6].

[3] Price quotations are an exception to the first part of the statement but usually not to the second.

equation models are usually models concerning such averages.

If then, one accepts the view that reactions between economic variables really take place in a short but non-zero interval, say the "reaction interval," whereas observations are on averages over a longer interval, say the "observation interval," what are the consequences as the reaction interval goes to zero and the observation interval remains fixed? The consequences for estimation remain to be worked out, so far as I am aware, although it seems likely that for some appropriate assumptions on the process generating the disturbances, the usual simultaneous equation estimators will emerge in the limit.

In the present paper, however, we are not concerned with estimation directly but with a rather different problem. We shall let the reaction interval go to zero, for fixed observation interval and for fixed values of the disturbances and exogenous variables and ask for necessary and sufficient conditions that the relations among the endogenous variables generate the same relations among their averages in the limit, save that the generated relations are to be simultaneous. We shall then ask what those conditions imply about the convergence of certain iterative methods of solving the model to obtain predicted values of the endogenous variables.

We are thus using the fact that the static structure assumed to relate current averages of endogenous variables must be generated by a dynamic process to obtain restrictions on that static structure. This is an application of Samuelson's "Correspondence Principle"[1] to econometrics.

To put the matter slightly differently, considerations of continuity

---

[1] Samuelson [9].

lead one to believe that models with extremely short time lags should yield approximately the same behavior as models with no time lags.[1] We examine the conditions that this should be true when, in both models, the observations are onaverages of the variables involved in the true relationships.

Thus, suppose that at any time, t, the simultaneous version of the model is given by the vector equations:

$$y(t) = G(y(t), z(t), u(t)) \tag{2.1}$$

where y(t) is an M-component vector of endogenous variables; z(t) is Λ-component vector of predetermined variables; and u(t) is an M-component vector of disturbances. G is (after estimation) a known vector-valued function. Identities have been substituted out of the model.[2]

We divide the unit time interval into n equal small subintervals, each of length $\Delta\theta = 1/n$.[3] Since we want to preserve the notation y(t) for the value of y at the instant t, we shall rewrite (2.1) as:

$$\bar{y}(t) = G(\bar{y}(t), z(t), u(t)), \tag{2.2}$$

---

[1] I am indebted to James S. Duesenberry for the perceptive suggestion (which prompted this study) that since very short time lags ought not to make much difference, simulation might well proceed as though such lags existed. The sense in which this is true is developed in the remaining sections.

[2] Identities must be substituted out because they obviously do hold instantaneously. If there are other equations of the model which are somehow believed to be truly instantaneous, they should be substituted out also before applying the procedures developed below. For convenience of exposition, we shall assume either that there are no such equations or that such substitution has already been made.

[3] Equality of the subintervals is a matter of convenience only.

where

$$(2.3) \qquad \bar{y}(t) \equiv \sum_{k=1}^{n} y(t + k\Delta\theta)\Delta\theta = \frac{1}{n} \sum_{k=1}^{n} y(t + k\Delta\theta).$$

We are interested in what happens when $\Delta\theta$ approaches zero,[1] when the true model is in fact not (2.1) or (2.2) but

$$(2.4) \qquad y(t + k\Delta\theta) = G\{y(t + (k-1)\Delta\theta), z(t), u(t)\} \qquad (k = 1, \ldots, n).$$

We shall ask what are the necessary conditions on G that (2.2) hold in the limit. Clearly, these will be some sort of stability conditions.

Now, it will be noticed that we have assumed the true model to have only a single lag. In fact, of course, there is no reason why this must be the case. The simultaneous model, (2.1) or (2.2), in a particular case might very well be the limit of a distributed lag model as the length of all lags approach zero. Moreover, this might be true even if the tests derived below showed that (2.2) cannot come from a single lag model such as (2.4). Indeed, it is generally so that one can invent <u>some</u> distributed lag model which generates (2.2) in the limit, provided one is not restricted to models in which the effect of a particular variable has the same sign for all lags.[2]

The justification for looking at only a single lag model is as follows.

---

[1] Note that when this occurs, $\bar{y}(t)$ approaches $\int_0^1 y(t + \theta)d\theta$. We shall not use this fact in the present paper.

[2] In the case of a linear model, it is known that "bunching" or "smearing" of effects does not affect stability if all effects are nonnegative or in certain other cases. See Bear [1].

Suppose that a particular simultaneous model fails our tests. Then that model can be the limiting case of some nonsimultaneous model but not of a single-lag model such as (2.4). To rely on this, however, is to take the position that small lags are negligible, but only if they are unequal and, indeed, only if they are unequal in certain specific ways. This is a possible position, but not a comfortable one.

To put it another way, if one is to ignore small lags, one wants to be able to ignore them no matter how they are distributed. One wants the simultaneous approximation to be appropriate no matter how effects are "bunched" or "smeared" over very small time intervals, provided some reasonable continuity of behavior is preserved.[1] Since it is generally not possible to examine conditions in general distributed lag models, we must content ourselves with examining single lag models in which all effects are bunched into a single time period. Necessary conditions that the simultaneous model be approached by such a single lag model will obviously also be necessary conditions that the exact lag structure not matter to the simultaneous equation approximation.[2] Those conditions are thus clearly necessary if one is to take the position that simultaneous models are limiting cases of non-simultaneous ones in which very short lags and their distribution have been neglected.

---

[1] One would hardly care if one's simultaneous model could not come from a distributed lag model with the same variable having wildly different effects at two very close points in time.

[2] Under some additional strong restrictions, they will also be sufficient. See Bear [1].

3. The Full Model: Linear Case

We begin with the case in which the model is linear, as it is here very easy to see what is involved and as useful results can be proved which are stronger than those available in the general case.

Since we are taking G linear, the true model generating the observations may be written as:

$$y(t + k\Delta\theta) = Ay(t + (k-1)\Delta\theta) + b \qquad (k = 1, \ldots, n) \tag{3.1}$$

where A is a constant M × M matrix and b is a constant M-component vector embodying the effects of z(t) and u(t) which we shall assume constant over the observation period.[1]

Substituting (3.1) into (2.3), we obtain:

$$\begin{aligned} \bar{y}(t) &= \sum_{k=1}^{n} y(t + k\Delta\theta)\Delta\theta = A \sum_{k=0}^{n-1} y(t + k\Delta\theta)\Delta\theta + nb\Delta\theta \\ &= A\bar{y}(t) + b + A\{y(t) - y(t + n\Delta\theta)\}\Delta\theta \\ &= A\bar{y}(t) + b + A\{y(t) - y(t+1)\}\Delta\theta. \end{aligned} \tag{3.2}$$

---

[1] This appears a strong assumption. We make it for two reasons. First, we want to concentrate on the internal dynamics of the endogenous variables. Second, in the absence of this assumption, it will be rare that simultaneous approximations in the averages will ever be appropriate, so that some such assumption is implicitly made in simultaneous models. In effect, what is involved is assuming that the effects of predetermined variables and random shocks are spread evenly over the observation period. Put this way, the assumption obviously is generally made in any model not dipping into the internal affairs of an observation period, which is the usual state of affairs. Otherwise complexities similar to those arising in solutions of mixed difference-differential equations can occur.

It is clear that this approaches the simultaneous version:

$$\bar{y}(t) = A\bar{y}(t) + b \tag{3.3}$$

in the limit if and only if:

$$\lim_{\Delta\theta \to 0} \{y(t) - y(t + n\Delta\theta)\}\Delta\theta = 0. \tag{3.4}$$

Since $y(t)$ is independent of $\Delta\theta$, this only involves the behavior of $y(t + n\Delta\theta)\Delta\theta$.

Since the simultaneous version (3.3) is supposed to be a theory of the determination of $\bar{y}(t)$, we may assume $(I - A)$ nonsingular.[1] Suppose that the characteristic values of A are all distinct. Then the solution of the difference equation (3.1) implies:

$$y(t + n\Delta\theta) = (I - A)^{-1}b + \sum_{j=1}^{M} K_j \lambda_j^n \eta_j \tag{3.5}$$

where the λ_j are the characteristic values of A, the η_j are the corresponding characteristic vectors, and the K_j are constants depending on $y(t)$.

For (3.4) to be satisfied no matter what $y(t)$ happens to be (that is, for initial conditions to wash out), it is thus necessary and sufficient in the present case that:

---

[1] This assumption is invariably made in simultaneous equation models. As it happens, it can be derived in the present case from slightly different considerations. If $(I - A)$ were singular, the solution of the difference equation (3.1) for $y(t + n\Delta\theta)$ would involve a term going to infinity at least as fast as n, and (3.4) could not hold.

$$(3.6) \qquad \lim_{\Delta\theta \to 0} \lambda_j^n \Delta\theta = \lim_{n \to \infty} \frac{\lambda_j^n}{n} = 0 \qquad (j = 1, \ldots, M).$$

This implies:

$$(3.7) \qquad |\lambda_j| \leq 1 \qquad (j = 1, \ldots, M)$$

where $|\lambda_j|$ denotes the modulus of λ_i. It is also easy to show that this condition is implied even if A has multiple roots.

Moreover, since (I - A) would be singular if unity were a characteristic value of A, we have shown that a necessary and sufficient condition for the validity of the simultaneous approximation in the limit in the linear case is:

$$(3.8) \qquad |\lambda_j| \leq 1, \qquad \lambda_j \neq 1 \qquad (j = 1, \ldots, M).$$

4. Digression: Solution by Iteration in the Linear Case

We may briefly point out that if the conditions (3.8) are satisfied, then solution of the system for the values of the endogenous variables can be accomplished by a simple iterative process. This is of only moderate importance for the linear case, but its extension to the nonlinear case (below) is important and it is thus worth taking up briefly.

Define:

$$(4.1) \qquad y^* = (I - A)^{-1} b$$

and for any scalar parameter α, $0 < \alpha \leq 1$, consider the iteration:

$$(4.2) \qquad y_{i+1} = \alpha(Ay_i + b) + (1 - \alpha)y_i = (\alpha A + (1 - \alpha)I)y_i + \alpha b.$$

It is easy to show that, if this converges, it converges to y^*. Further, it will converge if and only if all characteristic values of $B \equiv (\alpha A + (1 - \alpha)I)$ are less than one in modulus.

Now, let ρ_j be a characteristic value of B. There then exists a λ_j (a characteristic value of A), such that:

$$(4.3) \qquad \rho_j = \alpha\lambda_j + (1 - \alpha) \qquad (j = 1, \ldots, M).$$

Write λ_j as $c_j + d_j i$ where i is the square root of minus one and c_j and d_j are real. Then (3.8) implies:

$$(4.4) \qquad c_j^2 + d_j^2 \leq 1; \qquad c_j < 1 \qquad (j = 1, \ldots, M).$$

From (4.3) and (4.4), for any α <u>strictly</u> between zero and one:

$$(4.5) \qquad |\rho_j|^2 = \{\alpha c_j + (1 - \alpha)\}^2 + (\alpha d_j)^2 = \alpha^2(c_j^2 + d_j^2) + (1 - \alpha)^2 + 2\alpha c_j(1 - \alpha)$$

$$< \alpha^2 + (1 - \alpha)^2 + 2\alpha(1 - \alpha) = \{\alpha + (1 - \alpha)\}^2 = 1 \qquad (j = 1, \ldots, M).$$

We have thus shown that, in the linear case, the validity of the simultaneous approximation implies the convergence of the iteration (4.2) for any α chosen with $0 < \alpha < 1$. Of course, convergence will frequently also occur with $\alpha = 1$, but this is not guaranteed. Clearly, also, the choice of α will typically affect the speed of convergence.

5. The Full Model: General Case

In this section, we generalize the results for the linear case so far as possible, although the results obtained are necessarily weaker because of the impossibility of writing a general solution to general nonlinear difference equations. (Note, incidentally, that our results for the linear case do not carry over directly into local results for the nonlinear case. A condition such as (3.4) does not have to hold locally to be valid in the large.)

Let y^* be a fixed point of the mapping (2.2); that is,

$$y^* = G(y^*, z(t), u(t)) \equiv g(y^*) \tag{5.1}$$

where the second equality defines $g(y^*)$. Since (2.2) is supposed to provide a theory of the determination of $\bar{y}(t)$, we shall assume that for any given values of $z(t)$ and $u(t)$, y^* exists and is unique (although its value of course depends on those of $z(t)$ and $u(t)$).[1]

Next, define, for any M-component vector, x:

$$g^k(x) \equiv g\{g^{k-1}(x)\}; \quad g^0(x) \equiv x \qquad (k = 1, \ldots, \text{ad inf.}). \tag{5.2}$$

The underlying true relations to which (2.2) is the simultaneous

---

[1] The assumption of uniqueness can be relaxed to the assumption that all but one y^* can be ruled out by prior considerations. This would suffice for our purposes, but would slightly complicate the discussion in the text. Note that the assumption is the natural extension of that of the nonsingularity of (I - A) in the linear case. Cf. Christ [3, p. 213].

approximation are:

$$(5.3) \qquad y(t + k\Delta\theta) = g\{y(t + (k-1)\Delta\theta)\} \equiv G\{y(t + (k-1)\Delta\theta), z(t), u(t)\}$$

$$(k = 1, \ldots, n).$$

It is clear that the simultaneous approximation will be valid in the limit if and only if:

$$(5.4) \qquad \lim_{\Delta\theta\to 0} \bar{y}(t) \equiv \lim_{n\to\infty} \bar{y}(t) = y^*,$$

for any initial y(t). In other words:

$$(5.5) \qquad \lim_{n\to\infty}\left\{\frac{1}{n}\sum_{k=1}^{n} g^k(y)\right\} = y^*$$

for every y.[1] Obviously, for any finite y, we can also write:

$$(5.6) \qquad \lim_{n\to\infty}\left\{\frac{1}{n+1}\sum_{k=0}^{n} g^k(y)\right\} = y^*$$

for every y.[2]

---

[1] If some kind of regular behavior is imposed on z(t) and u(t), it may suffice to require that (5.5) hold only for y = y(t) in some appropriately defined neighborhood of y^* (see the preceding footnote). In this case, our results below hold also only in such a neighborhood.

[2] If there exists a $\hat{y}$, such that $y = g^j(\hat{y})$ for some finite integer j, then (5.6) follows directly from (5.5) without appeal to the fact that (y/n+1) approaches zero. In general, this will be the case.

Since a sum can converge only if its last term approaches zero, this means that for any y, $g^n(y)$ can at most go to infinity slower than n. In the linear case, this led immediately to the conclusion that $g^n(y)$ was at least bounded if not convergent (see (3.7)), but this does not follow in the general case. For the general case, (5.6) is itself the necessary and sufficient condition that simultaneity be the limiting approximation of the true system.

6. Digression: Solution by Iteration in the General Case

Now suppose that (5.6) holds and consider the following class of iterative schemes. Starting at an arbitrary point, y_0, choose some integer $n \geq 1$, and define:

$$y_{i+1} = \frac{1}{n+1} \sum_{k=0}^{n} g^k(y_i) \qquad i = 0, 1, \ldots, \text{ad inf.} \tag{6.1}$$

In view of (5.6), for any given y_0 and any arbitrary $\varepsilon > 0$, there exists an n sufficiently large that, defining y_1 by (6.1),

$$|y_1 - y^*| \leq \varepsilon |y_0 - y^*| \tag{6.2}$$

where $|x|$ denotes the length of x, for any vector x. Furthermore, if y_0 is restricted to lie in a compact set, then the requisite n can be chosen independent of y_0.

Now, for any y_0, consider the compact set $S(y_0) = \{y : |y - y^*| \leq |y_0 - y^*|\}$. Choose $\varepsilon < 1$. Choose n sufficiently large to make (6.2) hold with y_0 replaced by any $y \in S(y_0)$. Now perform the iteration (6.1). By (6.2), y_1 is in $S(y_0)$, whence repeated application of (6.2) shows $y_i \in S(y_0)$,

$i = 0, 1, \ldots,$ ad inf. Moreover,

$$|y_i - y^*| \leq \varepsilon^i |y_0 - y^*| \qquad (i = 1, \ldots, \text{ad inf.}) \tag{6.3}$$

so the iteration converges to y^*.

While the n chosen in the above depends on y_0, in practice y_0 can always be taken to lie in a bounded set (since the number of values tried in actual iteration will be finite), so there is some n which will work for all actual y_0.

Thus an iteration based on an unweighted average of the $g^k(y)$ converges to y^*. What about an iteration using a weighted average? We saw above that in the linear case, the iteration:

$$y_{i+1} = \alpha g^k(y_i) + (1-\alpha) y_i \tag{6.4}$$

was guaranteed to converge for every choice of α with $0 < \alpha < 1$, provided our conditions were satisfied.[1] For $\alpha = \frac{1}{2}$, this is equivalent to the iteration (6.1) with n = 1. Now consider the weighted average:

$$\hat{y} = \sum_{k=0}^{n} w_k(n) g^k(y), \tag{6.5}$$

where

$$w_k(n) \geq 0, \qquad \sum_{k=0}^{n} w_k(n) = 1. \tag{6.6}$$

---

[1] This form of iteration is a well-known method of numerical analysis. It is the Jacobi iterative method with a damping factor.

Defining $\bar{y}$ as the unweighted average of $g^0(y), \ldots, g^n(y)$, we have:

$$(6.7) \qquad \hat{y} - \bar{y} = \sum_{k=0}^{n} \left\{ w_k(n) - \frac{1}{n+1} \right\} g^k(y).$$

In view of (5.6), $\bar{y}$ approaches y^* as n becomes infinite. Further, we have already observed that:

$$(6.8) \qquad \lim_{n\to\infty} \frac{g^n(y)}{n+1} = 0.$$

It follows that provided that $w_k(n)$ approaches $(1/n+1)$ sufficiently fast for all $k = 1, \ldots, n$ as n becomes infinite, $\hat{y}$ will also approach y^*.

Of course, such a condition is not necessary. In view of the fact that (5.6) must hold for any initial y, it will suffice that $w_k(n)$ approach zero for $k = 0, \ldots, h$, with h some finite number and $w_k(n)$ approach $(1/n-h)$ sufficiently fast for $k = h+1, \ldots,$ ad inf. Further, if $g^n(y)$ itself approaches y^*, not even this is necessary.

Now, if $\hat{y}$ does in fact approach y^* for all initial y, choose an $n \geq 1$, and consider the iteration:

$$(6.9) \qquad y_{i+1} = \sum_{k=0}^{n} w_k(n) g^k(y_i) \qquad (i = 0, 1, \ldots, \text{ad inf.}).$$

By an argument identical to that given for the convergence of (6.1) above, for large enough n, this will converge to y^* for any y_0 in a bounded set.

Consider then, weighted averages in the form (for example):

$$(6.10) \qquad w_k(n) = \frac{1}{n+1} + \left(\frac{d}{k}\right) f(n) \qquad (k = 1, \ldots, n)$$

$$w_0(n) = 1 - \sum_{k=1}^{n} w_k(n)$$

for d some suitably chosen constant and f(n) a scalar function which approaches zero as n goes to infinity and does so sufficiently fast as to make $\hat{y}$ converge to y^*. Choose an α between zero and one, set n = 1, and $df(1) = \alpha - 1/2$. We obtain the iteration (6.4).

Thus, for the general case, we have shown that a generalized form of the iteration (6.4) converges if the simultaneous approximation is a valid one. We have not shown that for some α that iteration itself must converge (and, a fortiori, not that it must converge for all α between zero and one, as occurs in the linear case), but, of course, appropriate choice of α may in fact lead to convergence.[1] In any case, the unweighted average of the $g^k(y)$ must so converge, and this is very easy to check computationally (for example, by the iteration suggested in (6.1)).

7. The Results Strongly Stated: Normalization Rules and Submodels

What have we shown so far? If simultaneous equation models are considered as averaged approximations to nonsimultaneous models in the way described, we have shown that a necessary property of such models is the convergence of the sum (5.6), for given values of z(t) and u(t), to the values of the endogenous variables required for forecasting or simulation purposes. Further, if the model is entirely linear, we obtain a condition

---

[1] In general, one rather expects that it will, particularly in view of the results for the linear case. Since the unweighted average of the $g^k(y)$ converges, one suspects that an artful choice of a weighted average of a particular $g^k(y)$ and $g^{k+1}(y)$ will yield a vector closer to y^* than $g^k(y)$ itself, and that it may be possible to continue this. I have been unable to prove this, however.

on the characteristic roots of the matrix of coefficients of the current endogenous variables.

We can go considerably beyond this in two respects, however. The form in which (2.1) is written requires a choice of normalization rules, such that each equation of the model contains a different current endogenous variable on the left-hand side and such that each current endogenous variable appears on the left-hand side of exactly one equation. It is clear that the choice of normalization rules involved in writing (2.1) is not unique. Moreover, it is perfectly possible that the unweighted average (5.6) converges for one choice of normalization rules and fails to converge for another. If all that were involved in our results were experimentation with algorithms for simulation, we might experiment with different choices of normalization rules hoping to find one such choice for which (5.6) held.

Our results are stronger than this, however. We did not merely find that a properly specified simultaneous model must have the property that for some choice of normalization rules the unweighted average (5.6) converges. We showed that such iteration must converge in such a model for the particular choice of normalization rules corresponding to the true, nonsimultaneous system (5.3) which is being approximated. There is no liberty as to the choice of normalization rules in the equations of the true system[1] and none in the application of our results.

Another way of putting this is to observe that, in practice, every equation in a simultaneous system has a natural normalization rule (that of the corresponding equation of the true, nonsimultaneous system). The

---

[1]Recall that identities have been substituted out.

consumption function has consumption on the left, the investment demand function has investment demand, and so forth. Every equation of the true model represents the behavior of decision makers who set the value of a particular endogenous variable in response to the stimuli provided by their perceptions of the values of other variables a moment ago. Apparent ambiguity as to the natural normalization rule in a particular equation stems not from the simultaneity of the system but from ignorance of the nature of the true dynamic process whose static approximation is the simultaneous system considered. In practice, the natural normalization rules for most equations are perfectly obvious, as the examples just given indicate.[1] Our results apply with the equations written in their naturally normalized forms.[2]

If the fact that our results are required to hold for a particular choice of normalization rules imposes a strong condition on well-specified econometric models, the considerations to which we now turn impose much stronger conditions. Moreover, those conditions are directly useful in locating specification errors.

Simulation experiments with large models frequently take the following form.[3] The behavior of part of the model is to be investigated, so the remainder of the model is suppressed. This is done by dropping some subset

---

[1]The only serious exceptions to this take the form of the classic ambiguity as to whether price or quantity should appear on the left-hand side of a supply or demand equation. That ambiguity reflects the highly unsatisfactory state of our knowledge of the way in which prices are formed and quantities demanded or supplied in disequilibrium.

[2]Whether the naturally normalized forms are the best ones to use for estimation, however, depends on other considerations, principally the effect on the properties of the estimators used. Practically nothing is known about this issue. On this problem, see Fisher [5] and Mitchell and Fisher [7].

[3]See, for example, De Leeuw [4].

of the equations and taking the variables appearing on the left-hand side of those equations as exogenous for the sake of the experiment. Such an experiment, which of course is designed to show how parts of the economy, as represented by the model, react when taken in isolation, correspond to hypothetical real world situations in which the government, say, steps in and controls the level of certain variables rather than letting them adjust as in the period over which the model was estimated.

Now consider such an experiment. In our notation, this amounts to removing one or more of the elements of $y(t)$ from the list of endogenous variables and adding them to the list of variables included in $z(t)$, at the same time removing the appropriate equations from (2.1) and changing the definition of G, accordingly. The resulting submodel is then treated as though it were complete, the moved variables being assumed given from outside.

The clear (and reasonable) implication of this, however, is that the equations of the submodel being tested are the same and have the same parameter values whether or not that submodel is included in the larger model. This means that the submodel must be considered a valid simultaneous approximation to the corresponding submodel of the true dynamic model (5.3). It is then clear that (5.6) must hold <u>for the submodel</u>. Note that this is not implied by the validity of (5.6) for the full model, for then the process generating successive $g^k(y)$ is not the same as it is when some elements of y and functons of g are deleted. It is an implication of the position that a model can be put together from separate equations.

That implication, however, is a very strong one. We have just shown that the following is a necessary condition that a simultaneous equation model be well-specified. Not only must (5.6) hold for the model as a whole,

but it must also do so for any submodel generated by deleting from the model one or more of the equations and treating the corresponding left-hand side variable as predetermined. The strength of this requirement may perhaps be best seen by noting that, in the linear case, it is equivalent to the requirement that the characteristic values of every principal submatrix of A obey (3.8).

There are in principle 2^M-1 submodels to be examined in this requirement, but at least M, and generally many more of them will either trivially meet it or need not be separately considered so that the computational burden involved may not be excessive, given the simplicity of computation of (5.6).[1] We shall have more to say on this below.

We have thus developed a set of very strong criteria which must be obeyed by any simultaneous equation model which is capable of being considered an averaged approximation to a true dynamic model with vanishingly short time lags.[2] Those criteria should be applied to every simultaneous

---

[1] $\sum_{j=0}^{M} \binom{M}{j} = (1 + 1)^M$. There is obviously no point in checking submodels which are not simultaneous. This means that the M single equations of the model need not be considered, nor should models consisting of essentially unrelated equations with no feedback loops. Further, there is no point in testing separately a submodel containing two or more blocks of endogenous variables with no instantaneous feedback relating them (i.e., for which the relevant coefficient matrix is block-triangular). Such a submodel has the required property if and only if its component parts do. (Note that even if the full matrix is not block-triangular, it may become so when one or more equations are suppressed.) The submodels which are involved in a nontrivial way are generally those which might plausibly be involved in simulation experiments such as those mentioned above.

[2] Of course, those criteria need only be obeyed for the true values of the parameters and (by continuity) for some set of points in the parameter space not a set of measure zero and close to the true parameters. Except in pathological cases where the true parameters just barely satisfy the criteria (for example, if the true matrix A in the linear case has a characteristic value with modulus exactly unity), those criteria will be satisfied for all

equation model immediately after it is estimated. If the model meets those criteria, then simulation and forecasting with the model will be computationally extremely easy.

8. But What If It Doesn't?

The case in which the model passes our tests is thus a happy one. The potential usefulness of our results, however, is much greater in the case of a model which fails so to pass them.

Suppose, then, that a particular submodel fails to have the property (5.6). According to our results this is not simply a matter of bad luck. There is something wrong with the submodel.[1] Moreover, the test that is failed in such a circumstance is a test of the entire submodel itself, rather than of its component equations taken separately. Each separate equation may appear entirely reasonable, fit the data well, and so forth, and yet if the submodel fails our test, something is wrong with at least one equation. So far as I am aware, this is the only internal (i.e., not forecasting) test which has this property. Since tests which do use forecasts are not very well developed, and since equation specifications tend to be chosen one at a time, this is an important property. Thus, if a forecasting test fails, the problem must be somewhere in the model, but where is hard to pin down. If our test

---

points in the parameter space sufficiently close to the true parameters. Sampling fluctuations may lead to violation of our criteria even if the true model satisfies it, however; we shall return to this in the following section.

[1] Provided of course that one agrees that simultaneous models are averaged approximations to nonsimultaneous ones. If the world really is simultaneous, none of this applies. Recall that equations believed to be truly instantaneously valid are treated as identities for our purposes and substituted out before applying our tests.

fails for a particular submodel, there is a problem in that submodel.

What, then, can such a problem be? There are two possibilities. Either every equation of the submodel is correctly specified, but sampling problems have led to a set of parameters for which (5.6) is violated, or else there is misspecification in at least one equation. In either case, the submodel with the particular parameter values estimated is not capable of being the simultaneous approximation to the corresponding dynamic submodel.

What should be done? Unfortunately, our results provide a tool of diagnosis, not of cure. However, some general remarks may be in order.

The failure of our test for a particular submodel should cause the investigator to think hard again about the specification of that submodel. If he is confident about that specification, he may prefer to think that the difficulty is caused by the particular set of parameters estimated from the sample. This view can fairly easily be verified (for small submodels) by seeing if changing the parameters by not unreasonable amounts leads to a position where our test is passed. (What should then be concluded about the true parameters is, however, a more difficult matter.) Since what is involved is the internal structure of the relationships among current endogenous variables, such tinkering need not involve all the parameters of the submodel.[1]

In the linear case, the inferential problem is that of testing whether

---

[1]Note, however, that misspecification of the predetermined parts of an equation of the submodel can perfectly well affect the internal dynamics of the endogenous variables in the model as estimated. If the problem is not simply a sampling one, therefore, attention must be given to the specification of entire equations, not just to the parts relating to current endogenous variables.

the charateristic roots of various matrices satisfy (3.8). As the matrices involved are not symmetric, I know of no way to do this analytically. Monte Carlo experiments (making use of the asymptotic distribution of the parameter estimates) could of course be performed in both the linear and nonlinear cases.

We now go on to the case in which it is decided that sampling fluctuations do not account for the failure of the model to pass our tests. In this case, something is wrong with the specification of the model.

If the problems of the model are not too severe and widespread, our results can be used to localize them. In the first place, it is wise to begin the testing with the smallest submodels. The fact that a given submodel does or does not pass our test carries, in general, no implication about whether a larger submodel including the given one will pass it.[1] Further, (5.6) can be satisfied for the full model without being so for every submodel. Nevertheless, if a given small submodel fails (5.6), some adjustment must be made in that submodel. This being so, there is no point in going on to test larger submodels including the given one until such adjustment has been made, since that adjustment may perfectly well affect the ability of the larger submodel to satisfy (5.6).

Moreover, suppose a submodel consisting of a particular R equations satisfies (5.6). Suppose, however, that many or all of the submodels obtained by deleting one of the R equations and substituting a particular

---

[1] Except for the case in which the larger submodel consists of two or more smaller ones with no feedback loop connecting them. As indicated in an earlier footnote, such block-triangular cases need not be tested separately.

R+1st equation fail to satisfy (5.6). It does not follow that the difficulty lies in the R+1st equation involved (because our results give necessary but not sufficient conditions for submodels to be healthy), but such a phenomenon will certainly arouse suspicion as to the location of the disease. A similar statement holds for submodels differing from each other by substitution of more than one equation. Assuming the original specification of the model to have been not too far off the mark in too many places, artful combination of economic reasoning with the battery of localized tests provided by our results can locate difficulties which might otherwise remain hidden.

9. Experience with an Actual Model[1]

To gain experience with the tests described above, we applied them (in part) to an extended version of the Klein-Goldberger model.[2] The model, its coefficients, and the values of the variables used as starting values are exhibited in the Appendix, Tables 1-4.

The model has 23 equations (excluding the identities), so the number of submodels which might potentially have to be tested is of the order of 2^{23}. In fact, the number which actually should be separately tested is far lower. In about half an hour of IBM 7094 time our program for finding submodels to be tested lists roughly 2,000 such submodels. From inspection of

---

[1]The computations here reported were performed at the M.I.T. Computation Center. Programming was done and computation supervised by J. Timothy McGettigan. This part of the research was supported by National Science Foundation Grant GS-1376.

[2]Slightly adapted, with kind permission, from Norman [8]. This particular model was used because of the availability of already written programs to do part of the work.

the results, it appears likely that there are a good many more, but probably not more than 4,000 at the very most. Since our interest does not center directly on this model, we did not attempt an exhaustive listing when there turned out to be so many.

It should be noted that a perfectly reasonable looking 12-equation model set up for test purposes has less than ten separately testable submodels. Clearly, the number depends intimately on the structure of the model as well as on the number of equations. Typically the number is relatively small if after the removal of zero or one or two equations the model breaks into a block-triangular structure.

Despite the high number of submodels encountered and the machine time required to find them, it is not at all infeasible to test them. With n = 50, the iteration (6.1) was tried (for the simple set of starting values given) for the full model and about 140 of the submodels. This took somewhat less than nine minutes of 7094 time.[1] Were our principal interest in the model rather than in the test, it would clearly have been possible to test all the submodels encountered without a prohibitive expenditure of machine time. As new-generation computers become available, it will become possible to do this for ever larger models, although it is impossible to say without experimenting how much machine time would presently be required to test fully a model such as the Brookings quarterly model. Even if machine time requirements are too high to perform tests on all submodels, it is still

---

[1] With n = 10, it took less than four minutes, but many of the submodels which converged for n = 50 diverged for n = 10 and did so very rapidly. The 9-minute figure given in the text is conservative because of an error affecting one case and consuming a nonnegligible amount of time.

Erratum:

A CORRESPONDENCE PRINCIPLE FOR SIMULTANEOUS EQUATION MODELS

by Franklin M. Fisher

Page 26, line 9: "annual earnings" should be "wage and salary workers"

possible (and eminently desirable) to test the full model and those submodels in whose separate functioning and economic sense one is particularly interested.

In the tests performed, convergence was obtained for the full model and for all but 34 of the submodels.[1] In general, convergence was extremely rapid, occurring on the first iteration. Divergence took the form of overflow of preset requirements and was also relatively rapid.[2]

The submodels for which the iteration diverged are listed in Appendix Table 5. It will be noted that equations 6 and 7 (corresponding endogenous variables: hours worked and annual earnings) form part of every such submodel. This tentatively suggests that these equations ought to be re-examined, although such a conclusion cannot be a strong one without the results of a complete set of tests. (The iteration did converge for approximately 50 submodels containing these two equations.) Note that the two equations are related in a way that makes economic sense.

In closing, we may note that the fact that the full model itself allows convergence means that the iteration (6.1) provides a computationally easy way to generate solution of the model for forecasting and simulation purposes. This is, of course, a general property for models passing our tests.

---

[1] Note that convergence of (6.1) does not imply that the limit point is y^*. This must be separately tested. No cases of convergence to other points were encountered in the tests.

[2] As stated in an earlier footnote, the test was also performed for n = 10. Here additional submodels diverged. There seems little doubt in the case of the 34 submodels diverging for n = 50 that the difficulty is intrinsic and not due to the choice of n, although some further experimentation would be required to make this absolutely certain.

References

[1] Bear, D. V. T., "Distributed Lags and Economic Theory," Review of Economic Studies, Vol. XXXIII (1966), pp. 235-243.

[2] Bentzel, R. and B. Hansen, "On Recursiveness and Interdependency in Economic Models," Review of Economic Studies, Vol. XXII (1954-55), pp. 153-168.

[3] Christ, C., Econometric Models and Methods, New York: John Wiley & Sons, 1966.

[4] De Leeuw, F., "Financial Markets in Business Cycles: A Simulation Study," American Economic Review, Vol. LIV (Papers and Proceedings, 1964), pp. 309-323.

[5] Fisher, F. M., "Dynamic Structure and Estimation in Economy-wide Econometric Models," Ch. 15 in J. S. Duesenberry et al., eds., The Brookings Quarterly Econometric Model of the United States, Chicago and Amsterdam: Rand McNally & Co. and North-Holland Publishing Co., 1965.

[6] Gorman, W. M., "Professor Strotz on a Specification Error," University of Birmingham, Faculty of Commerce and Social Science, Discussion Papers, Series A, No. 24, 1960 (unpublished).

[7] Mitchell, B. and F. M. Fisher, "The Choice of Instrumental Variables in the Estimation of Economy-wide Econometric Models: Some Further Thoughts" (forthcoming).

[8] Norman, M., "Solving a Nonlinear Econometric Model by the Gauss-Seidel Iterative Method," paper presented at Washington Meetings of the Econometric Society, December 1967.

[9] Samuelson, P. A., Foundations of Economic Analysis, Cambridge: Harvard University Press, 1953.

[10] Strotz, R., "Interdependence as a Specification Error," Econometrica, Vol. 28 (1960), pp. 428-442.

[11] Wold, H., in association with L. Juréen, Demand Analysis, New York: John Wiley & Sons, 1953.

Table 1

DEFINITION OF VARIABLES USED IN MODEL

1	C_d	Consumption of durables, billions of 1954 dollars.
2	C_n	Consumption of nondurables and services, billions of 1954 dollars.
3	R	Residential construction, billions of 1954 dollars.
4	H	Stock of inventories, billions of 1954 dollars.
5	I_m	Imports, billions of 1954 dollars.
6	h	Index of hours worked per week, 1954 = 1.00.
7	N_w	Wage and salary workers, millions.
8	W	Wages and salaries and supplements to wages and salaries, billions of 1954 dollars.
9	w	Annual earnings, thousands of dollars.
10	S_c	Corporate saving, billions of 1954 dollars.
11	P_c	Corporate profits, billions of 1954 dollars.
12	Π_r	Rental income and net interest, billions of 1954 dollars.
13	X	Gross national product, billions of 1954 dollars.
14	Y	Personal disposable income, billions of 1954 dollars.
15	Π	Proprietors' income, billions of 1954 dollars plus P_c and IVA.
16	IT	Indirect taxes, billions of current dollars.
17	T_c	Corporate profits taxes, billions of current dollars.
18	PT	Personal taxes, billions of current dollars.
19	SI	Contributions for social insurance, billions of current dollars.
20	BT	Business transfers, billions of current dollars.
21	GT	Government transfers, billions of current dollars.

[cont'd]

22	p	Implicit GNP deflator, 1954 = 1.00.
23	IVA	Inventory valuations adjustment, billions of current dollars.
24	I	Investment in plant and equipment, billions of 1954 dollars.
25	r_s	Yield on prime commercial paper, 4-6 months, per cent.
26	D	Capital consumption allowances, billions of current dollars.
27	r	Average yield on corporate bonds (Moody's), per cent.

Exogenous Variables

E1	ID	Net interest paid by government, billions of current dollars.
E2	W_g	Government wages and salaries, billions of 1954 dollars.
E3	p_m	Implicit price deflator for imports, 1954 = 1.00.
E4	N_g	Government employees, millions.
E5	N_s	Self-employed workers, millions.
E6	N_L	Total labor force, millions.
E7	GS	Subsides - current surplus of government enterprise, billions of current dollars.
E8	D_u	Dummy variable, 0 for 1929-1946, 1 for 1947-1962.
E9	r_d	Average discount rate at all Federal Reserve Banks, per cent.
E10	R_e	Year-end ratio of member banks' excess to required reserves.
E11	G	Government expenditures, billions of 1954 dollars.
	E	Exports, billions of 1954 dollars.
E12	SD	Statistical discrepancy, billions of current dollars.
E13	I_R	Value of last twenty years investment.
E14	WB	Average weekly benefits for unemployed.

Table 2

EQUATIONS OF THE MODEL

1 $Cd - a_1 Y = (.7 + a_2) Cd_{-1} - .7 a_1 Y_{-1} + a_3$

2 $Cn - a_4 Y = a_5 Cn_{-1} + a_6$

3 $R - a_{11} Y = a_{12} r_{-1} + a_{13} R_{-1} + a_{14}$

4 $(1 + a_{15})H - a_{15} X = (a_{15} + a_{16}) H_{-1} + a_{17}$

5 $Im - a_{18} X + a_{19} p = a_{19} Pm + a_{20} Im_{-1} + a_{21}$

6 $X - a_{22} R - a_{23} \frac{Nw}{9} - a_{24} h = Wg + .95(X - Wg)_{-1} + a_{22} I$

$- a_{23} Ng - Ns + .95(Nw - Ng + Ns)_{-1} - .95 a_{24} h_{-1} + a_{25}$

7 $h - a_{26} w + a_{27} Nw = - a_{26} w_{-1} + a_{27} (N1 - Ns) + a_{28}$

8 $W - a_{29} X = (1 - a_{29}) Wg + a_{30} (W - Wg)_{-1} + a_{31}$

9 $w + a_{32} Nw = w_{-1} + a_{32}(N1 - Ns) + a_{33} (P_{-1} - P_{-2}) + a_{34}$

10 $pSc + IVA - a_{38} (pPc + IVA - Tc) = a_{39} (pPc - Tc - pSc)_{-1} + a_{40}$

11 $p\Pi - pPc - IVA - a_{41} p X = a_{42} (p(\Pi - Pc) - IVA)_{-1} + a_{43}$

12 $p\Pi r - a_{44} pR - a_{44} pI = a_{45} (r - r_{-1}) + a_{46} (p\Pi r)_{-1} + a_{47}$

13 $X - Cd - Cn - R - H + Im = H_{-1} + I + \frac{G + E}{E_{11}}$

14 $pY - pX + pSc + IT + Tc + PT + SI - GT + IVA = GS - SD - D + ID$

15 $p\Pi - pX + pW + p\Pi r + IT + BT = - D + GS - SD$

16 $IT - C_2 pX = C_1$

[cont'd]

17 $Tc - C_4\ pPc = C_3$

18 $(1 - C_6)\ PT - C_8\ pY = C_5$

19 $SI - C_8\ (pY + PT) = C_7$

20 $BT - a_{57}\ pX = a_{58}\ BT_{-1} + a_{59}$

21 $GT + Cg\ Nw = Cg\ (N1 - Ns) + C_{10}\ WB + C_{11}$

22 $pW - wh\ Nw = 0$

23 $IVA - a_{55}\ p = - a_{55}\ p_{-1} + a_{56}$

24 $I = a_7\ (X - Wg)_{-1} + a_8\ r_{-1} + (.95 + a_9)\ I_{-1} + a_{10}$

25 $rs = a_{51}\ rd + a_{52}\ Re_{-1} + a_{53}\ Du\ \&\ a_{54}$

26 $D = a_{48}\ \Sigma p(I + R) + a_{49}\ Du + a_{50}$

27 $r = a_{35}\ rs + a_{36}\ r_{-1} + a_{37}$

In each case the equation was solved for the correspondingly-numbered endogenous variable (see Table 1) before testing.

Table 3

VALUES OF COEFFICIENTS

	$a_1 - a_{28}$		$a_{29} - a_{59}$		$C_1 - C_{11}$
1	0.2314	29	0.4956	1	0.6470
2	-0.1042	30	0.1307	2	0.0061
3	-0.4618	31	-1.2517	3	0.0591
4	0.2502	32	-2.1800	4	0.0807
5	0.7226	33	6.7879	5	-0.0175
6	-0.1171	34	2.1602	6	0.0330
7	0.0656	35	0.1693	7	0.0200
8	-0.2113	36	0.8116	8	-0.0000
9	-0.5900	37	0.4015	9	0.0517
10	0.9319	38	0.9009	10	-0.0000
11	0.0468	39	-0.8893	11	0.0778
12	-0.0462	40	0.0024		
13	0.3983	41	0.0096		
14	-0.1228	42	0.9092		
15	0.1347	43	0.0627		
16	0.4050	44	0.0752		
17	-2.4301	45	-0.1084		
18	0.0331	46	0.9131		
19	-1.6649	47	-0.0471		
20	0.3477	48	0.0492		
21	-0.1214	49	0.8556		
22	0.3344	50	-0.1411		
23	2.2442	51	1.1452		
24	1.8814	52	-0.0082		
25	-0.5863	53	0.5331		
26	-0.4050	54	-0.5107		
27	-1.8349	55	-4.5050		
28	11.4307	56	0.0002		
		57	0.0017		
		58	0.6149		
		59	0.0066		

Table 4

VALUES OF VARIABLES USED FOR INITIAL CONDITIONS

Endogenous Variables		Present	Lagged
1	C_d	1.49	1.41
2	C_n	11.33	10.73
3	R	0.87	1.18
4	H	0.30	0
5	I	1.09	1.02
6	h	11.16	12.11
7	N_w	3.76	3.65
8	W	8.90	8.06
9	w	12.16	11.42
10	S_c	0.42	.43
11	P_c	1.67	1.65
12	Π_r	2.07	1.87
13	X	18.18	16.98
14	Y	14.48	13.10
15	Π	4.33	3.93
16	IT	0.70	0
17	T_c	0.14	.13
18	PT	0.26	0
19	SI	0.02	0
20	BT	0.06	.05
21	GT	0.09	0
22	p	.57	.58
23	IVA	0.05	-.04
24	I	2.33	2.10
25	r_s	5.85	0
26	D	0.86	0
27	r	5.21	5.09

[cont'd]

Exogenous Variables		Present	Lagged
1	ID	.10	.0
2	W_g	0.86	.82
3	P_m	0.573	0
4	N_g	0.35	.34
5	N_s	1.03	1.03
6	N_L	4.94	0
7	GS	-0.01	0
8	D_u	0.00	0
9	r_d	5.00	0
10	R_e	2.05	0
11	G+E	2.96	0
12	SD	0.03	0
13	I+R	18.47	17.53
14	WB	0.58	.58
15	P_{-2}	.576	

Table 5

SUBMODELS TESTED AND DIVERGING (n = 50)

(Numbers correspond to Tables 1 and 2)

2	4	6	7					
1	3	6	7	8				
1	6	7	8	10				
1	6	7	10	19				
1	2	4	5	6	7			
1	3	5	6	7	9			
1	4	5	6	7	8			
1	5	6	7	8	10			
1	5	6	7	10	19			
2	5	6	7	10	19			
2	6	7	8	10	19			
2	6	7	10	11	12			
3	5	6	7	8	19			
1	2	3	6	7	10	11		
1	2	6	7	8	10	11		
1	5	6	7	10	11	12		
2	3	5	6	7	10	11		
2	4	5	6	7	10	19		
2	4	6	7	10	11	12		
2	5	6	7	10	11	12		
3	5	6	7	8	10	11		
1	2	3	4	6	7	8	19	
1	2	4	6	7	8	10	11	
1	4	5	6	7	8	10	11	
2	3	5	6	7	8	9	10	
3	4	6	7	10	11	12	19	
1	2	3	5	6	7	8	10	11
1	2	3	5	6	7	10	11	12
1	2	4	5	6	7	8	10	19
2	3	4	5	6	7	10	11	19
2	3	4	6	7	8	9	10	11
2	3	4	6	7	8	10	11	12
2	4	5	6	7	10	11	12	19
3	4	5	6	7	8	10	11	19